



## **National Statistics Review of Data on Sheep Numbers at December in England.**

**Author:** Julian Groom  
Survey Design & Analysis Branch  
1-2 Peasholme Green  
York, YO1 7PX  
Tel. (01904) 455435  
[Julian.groom@defra.gsi.gov.uk](mailto:Julian.groom@defra.gsi.gov.uk)

**Date:** November 2010.

## Overview

Defra currently has two sources of data on sheep in December: the December Survey of Agriculture and the Sheep and Goat (SAG) Inventory. This review aims to establish if it is possible to have one single data source for sheep in future. This would yield internal efficiencies and savings, avoid duplication of effort in data collection and reduce the burden of paperwork on farmers caused by Defra's surveys.

## Overall Conclusions

- The SAG Inventory will be the source of data on sheep at December and questions about sheep will be removed from the December Survey form from 2010 onwards. Results will cover all sheep holdings in England.
- Survey Support Section (SSS) to give priority to cleaning data from the SAG Inventory so that the best quality data on sheep is available by the 15<sup>th</sup> of February each year when work on results starts.
- SSS and the Animal Health Agency (AHA) concentrate on improving and maintaining the quality of the registers used to select recipients of SAG forms. This will help to improve the response rate and the quality of the data.
- Data on sheep numbers will continue to be collected through the June Survey (for commercial holdings only (a)). However, this will be considered as part of our project to review the levels of data collection required from farmers in future. We will be in contact when thoughts on this are further developed.

## Introduction

The SAG Inventory was introduced in December 2005 and asks respondents to supply the number of sheep and/or goats that they have on the 1<sup>st</sup> of December in England. The AHA uses this information to report to the EU on the UK's compliance with sheep and goat registrations and identification requirements every August. Respondents have to give details of their animals both on their own holding and on any other holdings or common land where they graze their animals. This means there are two holding numbers for each record in the database- one for the number of the keeper's main holding (MCPH) and one for any other holding(s) where they graze animals, referred to as the local holding number. Numbers of animals were aggregated by MCPH in this study, as the local numbers were entered manually in an inconsistent format. We deemed the data to be unreliable for this purpose as a result. Records in the databases used in this study were checked for outliers and were corrected accordingly where necessary. Respondents have to provide total numbers of sheep in their breeding flock and other sheep and lambs.

The December Survey of Agriculture was introduced in 1951 to complement the June Census of Agriculture. It provides a half-yearly snapshot of agriculture by capturing crop plantings and numbers of livestock in winter. Its survey day also falls on the 1<sup>st</sup> of December. Validation checks are run as the returns are keyed into our database and any queries raised about entries on the forms at this stage of the process are resolved by colleagues in SSS. The results of the December Survey are published in the following March and they currently include figures for five different types of sheep (shown in the table below).

Table 1: Comparison of Items Collected.

	December Survey	SAG Inventory
<b>Sheep (over 1 year old):</b>	Breeding ewes and shearlings* put to the ram	Breeding sheep & lambs put to the ram
	Rams for service Other sheep over 1 year old	Other sheep (including lambs & rams)
<b>Lambs (under 1 year):</b>	Ewe lambs put to ram in yyyy (current year) Other lambs	
<b>Total sheep and lambs</b>		

\* A shearling is a young sheep which has been sheared for the first time.

(a) Commercial holdings are defined as those with significant levels of farming activity ((as defined by EU Farm Structure Survey Regulation EC 1166/2008). These significant levels of farming activity are classified as any holding with more than five hectares of agricultural land, one hectare of orchards, 0.5 hectares of vegetables or 0.1 hectares of protected crops, or more than 10 cows, 50 pigs, 20 sheep, 20 goats or 1,000 poultry.

The SAG Inventory is required under EU law (Regulation no. 1505/2006) to collect data on the “Total number of sheep and lambs” on every holding in December each year. However, national sheep numbers are also required under a separate regulation (EU Regulation 1165/2008) which requires Member States to provide statistics on “ewes and ewe lambs put to the ram” and “other sheep”, as at December each year. Therefore, by including these two questions on sheep numbers in the SAG Inventory, this will enable the UK to continue to meet both regulation requirements in future

## Data Analysis

The 2009/10 SAG Inventory had a relatively high response rate in England, as shown in the table below:

Table 2: Overall Response Rate to the 2009/10 SAG Inventory in England.

Number of forms despatched	56,735
Number of forms returned	44,762
Overall response rate	78.9%
Raising factor (to account for non-response)	1.27

Assuming the non-respondents are sheep keepers, we need to adjust the raw data from the SAG Inventory to account for non-responses. I have checked whether the non-responses are valid, i.e. they are genuinely sheep keepers and should be included in the population (see Annex A for details of this analysis). I concluded that out of the 1,842 keepers who had not returned their 2009 SAG Inventory forms but had later returned a December or June form, 1,251 (68%) of them had subsequently recorded nearly half a million sheep on their December or June forms. This represents substantial under-recording of sheep in the 2009/10 SAG Inventory and clearly justifies the need to raise the “raw” sheep figures from the Inventory. It helps to emphasise the importance of keeping the AHA’s database of keepers as up to date as possible, as a response rate of 91% was achieved in Scotland’s 2009 SAG Inventory following a register improvement exercise.

The raised (i.e. raw data multiplied by raising factor shown above) totals for sheep items in England from the 2009/10 SAG Inventory and final figures for sheep from the 2009 December Survey were as follows:

Table 3: Comparison of Figures from the 2009/10 SAG Inventory and the 2009 December Survey.

	(thousands)				
	SAG Inventory	December Survey	Difference (SAG – Dec)	Diff as % of SAG Figure	Dec ‘09 95% Conf. Interval
Breeding Sheep	6,203	6,202	1	0%	± 538
Other Sheep (including Rams and Lambs)	3,741	3,672	69	1.8%	± 415
Total Sheep	9,944	9,873	71	0.7%	± 679

Table 3 above makes it clear that the totals from the 2009 SAG Inventory and December Survey are very similar, and the differences fall well within the December Survey’s confidence intervals. As you can see in the following table, this is repeated at the regional level in England:

Table 4: Regional Numbers of Sheep and Goats from the 2009/10 SAG Inventory (Raised Data) and the 2009 December Survey.

Region	Total Sheep (thousands)		Difference as % of SAG Figures	Total as % of Eng Total	
	SAG Inv	Dec Survey		SAG Inv	Dec Survey
North East	1,221	1,159	5%	12.3%	11.7%
North West	1,875	1,893	-1%	18.9%	19.2%
Yorkshire and The Humber	1,382	1,375	1%	13.9%	13.9%
East Midlands	771	751	3%	7.7%	7.6%
West Midlands	1,471	1,434	2%	14.8%	14.5%
East of England	266	264	1%	2.7%	2.7%
South East & London	831	950	-14%	8.4%	9.6%
South West	2,128	2,047	4%	21.4%	20.7%
England	9,944	9,873	1%	100%	100%

In conclusion, the final two columns show that the distribution of the total flock is similar from both data sources. This indicates that there is no geographic distortion caused by the SAG Inventory and that the MCPHs are allocated to their correct regions.

Datasets for the SAG Inventories in England are available for the period 2006 to 2009 so it is possible to compare a time series of SAG data against the corresponding December Surveys, as shown in the table below:

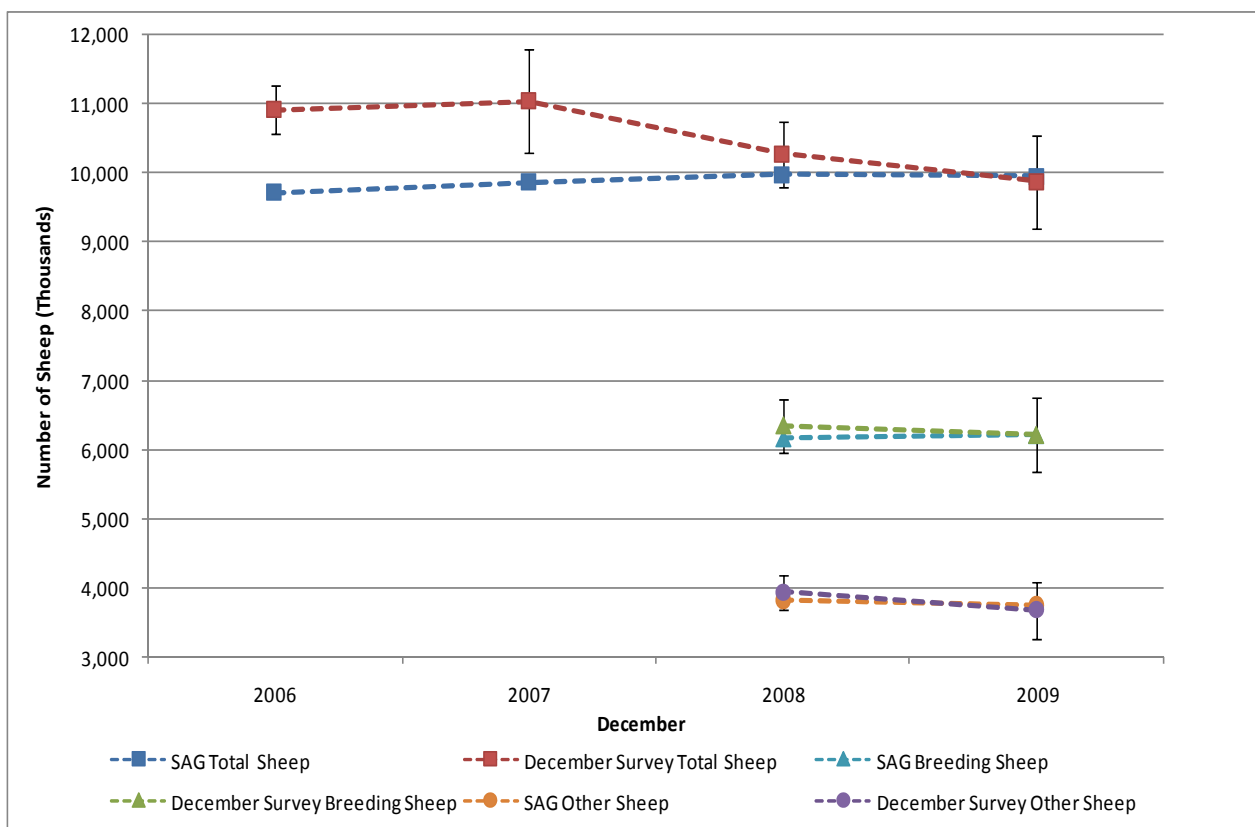
Table 5: Time Series of Sheep Data from the SAG Inventory and the December Survey, 2006 - 2009.

	2006		2007		2008		2009	
	SAG Inv	Dec Survey	SAG Inv	Dec Survey	SAG Inv	Dec Survey	SAG Inv	Dec Survey
Breeding Sheep	:	:	:	:	6,155	6,339	6,203	6,202
Other Sheep	:	:	:	:	3,814	3,929	3,741	3,672
Total Sheep	9,715	10,923	9,690	11,041	9,969	10,268	9,944	9,873

: = No comparison available. Only total sheep numbers were collected in 2006 and 2007 SAG Inventories.

Chart 1 below compares sheep figures from the two exercises in this time series, with the upper and lower limits of values from the December Survey (calculated using confidence intervals where available) represented as error bars on the graph:

Chart 1: Comparison of Sheep Data from the SAG Inventory and the December Survey, 2006 - 2009.



The divergence between the two sets of results for Total Sheep in the exercises held in 2006 and 2007 was due to relatively unreliable results for lambs captured in the SAG Inventories held in those years. Lambs were not mentioned specifically in those years' questions which resulted in under recording of lambs by many respondents. Lambs have been specifically mentioned in questions from 2008 onwards, resulting in much better recording of lambs. The overall coverage for these years has dramatically improved as a result. In conclusion, this chart illustrates very clearly the convergence in the figures from the SAG Inventory and the December Survey after 2007, which strengthens the argument for using the SAG Inventory as the source of data for sheep figures in the December Survey's results from 2010 onwards.

## Survey Timings

Under EU Regulation 1165/2008, sheep estimates for December have to be supplied to Eurostat by the 15<sup>th</sup> of February (provisional results) and the 15<sup>th</sup> of May (final results). I carried out an analysis to see if the SAG data from forms returned by the 15<sup>th</sup> of February is of suitable quality and accurate enough to use (see Annex A for details of this analysis). The following tables show the snapshots around these dates from the last two SAG Inventories:

Table 6: Situation in England on 18/02/10, 18/05/10 and at Closedown in 2009/10 SAG Inventory.

(sheep in thousands)

Date	Number of Forms Despatched	Number of forms received	Number of "clean" records	Total Sheep	Raising Factor	Raised Total Sheep
18/02/2010	56,735	43,757	42,565	7,375	1.30	9,563
18/05/2010	"	44,419	44,299	7,777	1.28	9,933
Closedown: 23/07/10	"	44,762	44,762	7,846	1.27	9,944

Table 7: Situation in England on 18/02/09, 18/05/09 and at Closedown in 2008/09 SAG Inventory.

(sheep in thousands)

Date	Number of Forms Despatched	Number of forms received	Number of "clean" records	Total Sheep	Raising Factor	Raised Total Sheep
18/02/2009	53,438	36,924	36,643	5,777	1.45	8,360
18/05/2009	"	41,972	41,949	7,780	1.27	9,905
Closedown: 13/07/09	"	41,978	41,978	7,831	1.27	9,969

The results show that estimates do change as more data is returned and records are cleaned. Therefore, we need to ensure we give a high priority to cleaning as much data as possible by the 15<sup>th</sup> of February in future, particularly for checks involving large changes in holdings' flock sizes as these are the most influential in determining the quality of the data. In conclusion, the time freed up by colleagues in SSS not having to validate sheep data in future December Surveys must be allocated to validating SAG Inventory forms instead.

## Users of Sheep data

Existing users of sheep data from the December Survey include Defra's Livestock Statistics branch, livestock policy teams, including the Livestock and Animal By-products team, and the RADAR team. Further afield, there are agricultural industry bodies such as EBLEX and the Meat and Livestock Commercial Services Ltd under the Agriculture and Horticulture Development Board's umbrella. As mentioned earlier, Defra also has to supply data on sheep to Eurostat.

The SAG Inventory is less detailed than the December Survey, as it collects numbers of two types of sheep, and the December Survey collects information on five types of sheep. There is a specific example of a potential gap in the data on numbers of "ewe lambs put to the ram", which are used when predicting future trends in the size of the breeding flock by industry. This is inevitable given the less detailed categories recorded in the inventory. There is no scope to expand these categories as they are bound by European requirements and we need to show a genuine commitment to reducing the survey burden on farmers.

A similar scenario was found when we switched cattle data sources from survey to the Cattle Tracing System (CTS) when data on "heifers in calf" was lost. However, the benefits gained from switching to the administrative sources outweigh the negatives in terms of better quality, more accurate data with high coverage.

## Conclusion

The findings of this feasibility study are favourable for using the SAG Inventory, rather than the December Survey, as the source of data on sheep as at December in future, and the results from the two data sources are comparable.

Similarities between the two sets of sheep figures for England from the 2009 SAG Inventory and December Survey are reinforced at the regional level of geography, given the similar regions' proportions of the totals for England for the two sources of data. Since 2008 when the questions on the SAG forms were improved by specifically mentioning lambs, there has been "convergence" in the figures from the SAG Inventory and the December Survey.

Nearly half a million sheep were recorded by keepers on December and June Survey forms after they had not returned their 2009 SAG Inventory forms. This represents substantial under-recording of sheep in the 2009/10 SAG Inventory and clearly justifies the need to raise the "raw" sheep figures from the Inventory. The AHA, in particular, needs to ensure that the register used to select recipients of SAG forms is kept up to date to ensure a very high response rate and to produce high quality data in future. Work by SSS on SAG Inventory forms must have high priority in the future to ensure that the sheep results can be produced on time and to a high quality.

These factors combine to strengthen the argument for using the SAG Inventory as the source of data for sheep figures in the December Survey's results from 2010 onwards.

To re-cap, here are the findings of this review:

- The SAG Inventory will be the source of data on sheep at December and questions about sheep will be removed from the December Survey form from 2010 onwards. Results will cover all sheep holding in England.
- Survey Support Section (SSS) to give priority to cleaning data from the SAG Inventory so that the best quality data on sheep is available by the 15<sup>th</sup> of February each year when work on results starts.
- SSS and the Animal Health Agency (AHA) concentrate on improving and maintaining the quality of the registers used to select recipients of SAG forms. This will help to improve the response rate and the quality of the data.
- Data on sheep numbers will continue to be collected through the June Survey (for commercial holdings only (a)). However, this will be considered as part of our project to review the levels of data collection required from farmers in future. We will be in contact when thoughts on this are further developed.

(a) Commercial holdings are defined as those with significant levels of farming activity (as defined by EU Farm Structure Survey Regulation EC 1166/2008). These significant levels of farming activity are classified as any holding with more than five hectares of agricultural land, one hectare of orchards, 0.5 hectares of vegetables or 0.1 hectares of protected crops, or more than 10 cows, 50 pigs, 20 sheep, 20 goats or 1,000 poultry.